10/578084
IAP12 Rec'd PCT/PTO 08 MAY 2006

Method for distinguishing AML subtypes with aberrant and prognostically intermediate karyotypes

The present invention is directed to a method for distinguishing AML subtypes with aberrant and prognostically intermediate karyotypes, with respect to their specific prognosis based on genes as detected by gene expression profiling.

Leukemias are classified into four different groups or types: acute myeloid (AML), acute lymphatic (ALL), chronic myeloid (CML) and chronic lymphatic leukemia (CLL). Within these groups, several subcategories can be identified further using a panel of standard techniques as described below. These different subcategories in leukemias are associated with varying clinical outcome and therefore are the basis for different treatment strategies. The importance of highly specific classification may be illustrated in detail further for the AML as a very heterogeneous group of diseases. Effort is aimed at identifying biological entities and to distinguish and classify subgroups of AML which are associated with a favorable, intermediate or unfavorable prognosis, respectively. In 1976, the FAB classification was proposed by the French-American-British co-operative group which was based on cytomorphology and cytochemistry in order to separate AML subgroups according to the morphological appearance of blasts in the blood and bone marrow. In addition, it was recognized that genetic abnormalities occurring in the leukemic blast had a major impact on the morphological picture and even more on the prognosis. So far, the karyotype of the leukemic blasts is the most important independent prognostic factor regarding response to therapy as well as survival.

Usually, a combination of methods is necessary to obtain the most important information in leukemia diagnostics: Analysis of the morphology and cytochemistry of bone marrow blasts and peripheral blood cells is necessary to establish the diagnosis. In some cases the addition of immunophenotyping is mandatory to separate very undifferentiated AML from acute lymphoblastic leukemia and CLL. Leukemia subtypes investigated can be diagnosed by cytomorphology alone, only if an expert reviews the smears. However, a genetic analysis based on chromosome analysis, fluorescence in situ hybridization or RT-PCR and immunophenotyping is required in order to assign all cases in to the right category. The aim of these techniques besides diagnosis is mainly to determine the prognosis of the

leukemia. A major disadvantage of these methods, however, is that viable cells are necessary as the cells for genetic analysis have to divide in vitro in order to obtain metaphases for the analysis. Another problem is the long time of 72 hours from receipt of the material in the laboratory to obtain the result. Furthermore, great experience in preparation of chromosomes and even more in analyzing the karyotypes is required to obtain the correct result in at least 90% of cases. Using these techniques in combination, hematological malignancies in a first approach are separated into chronic myeloid leukemia (CML), chronic lymphatic (CLL), acute lymphoblastic (ALL), and acute myeloid leukemia (AML). Within the latter three disease entities several prognostically relevant subtypes have been established. As a second approach this further sub-classification is based mainly on genetic abnormalities of the leukemic blasts and clearly is associated with different prognoses.

The sub-classification of leukemias becomes increasingly important to guide therapy. The development of new, specific drugs and treatment approaches requires the identification of specific subtypes that may benefit from a distinct therapeutic protocol and, thus, can improve outcome of distinct subsets of leukemia. For example, the new therapeutic drug (STI571) inhibits the CML specific chimeric tyrosine kinase BCR-ABL generated from the genetic defect observed in CML, the BCR-ABL-rearrangement due to the translocation between chromosomes 3 and 22 (t(9;22) (q34; q11)). In patients treated with this new drug, the therapy response is dramatically higher as compared to all other drugs that had been used so far. Another example is the subtype of acute myeloid leukemia AML M3 and its variant M3v both with karyotype t[15;17)(q22; q11-12). The introduction of a new drug (all-trans retinoic acid - ATRA) has improved the outcome in this subgroup of patient from about 50% to 85 % long-term survivors. As it is mandatory for these patients suffering from these specific leukemia subtypes to be identified as fast as possible so that the best therapy can be applied, diagnostics today must accomplish sub-classification with maximal precision. Not only for these subtypes but also for several other leukemia subtypes different treatment approaches could improve outcome. Therefore, rapid and precise identification of distinct leukemia subtypes is the future goal for diagnostics.

Thus, the technical problem underlying the present invention was to provide means for leukemia diagnostics which overcome at least some of the disadvantages of the prior art diagnostic methods, in particular encompassing the time-consuming and unreliable

combination of different methods and which provides a rapid assay to unambigously distinguish one AML subtype from another, e.g. by genetic analysis.

According to Golub et al. (Science, 1999, 286, 531-7), gene expression profiles can be used for class prediction and discriminating AML from ALL samples. However, for the analysis of acute leukemias the selection of the two different subgroups was performed using exclusively morphologic-phenotypical criteria. This was only descriptive and does not provide deeper insights into the pathogenesis or the underlying biology of the leukemia. The approach reproduces only very basic knowledge of cytomorphology and intends to differentiate classes. The data is not sufficient to predict prognostically relevant cytogenetic aberrations.

Furthermore, the international application WO-A 03/039443 discloses marker genes the expression levels of which are characteristic for certain leukemia, e.g. AML subtypes and additionally discloses methods for differentiating between the subtype of AML cells by determining the expression profile of the disclosed marker genes. However, WO-A 03/039443 does not provide guidance which set of distinct genes discriminate between two subtypes and, as such, can be routineously taken in order to distinguish one AML subtype from another.

The problem is solved by the present invention, which provides a method for distinguishing AML subtypes with aberrant and prognostically intermediate karyotypes selected from trisomy 8, inv(3), t(3;3), trisomy 11, trisomy 13, trisomy 4, t(1;3), t(6;9), der(5)t(5;11), i(17), del(9q), del(12p), and/or del(20q) into different subsets in a sample, the method comprising determining the expression level of markers selected from the markers identifiable by their Affymetrix Identification Numbers (affy id) as defined in Tables 1 and/or 2,

wherein

a higher expression of at least one polynucleotide defined by any of the numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 42, 43, 44, 46, 47, 48, 49, and/or 50 of Table 1, and/or

a lower expression of at least one polynucleotide defined by any of the numbers 18, 41, and/or 45 of Table 1,

is indicative for a specific median event-free survival (EFS) and

and/or wherein

a higher expression of at least one polynucleotide defined by any of the numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 45, 49, and/or 50 of Table 2, and/or

a lower expression of at least one polynucleotide defined by any of the numbers 24, 44, 46, 47, and/or 48, of Table 2

is indicative for a specific median overall survival (OS),

As used herein,

trisomy 8 means AML with trisomy of chromosome 8

inv(3) means AML with inversion 3

t(3;3) means AML with translocation t(3;3)

trisomy 11 means AML with trisomy of chromosome 11

trisomy 13, means AML with trisomy of chromosome 13

trisomy 4 means AML with trisomy of chromosome 4

t(1;3) means AML with translocation (t1;3)

t(6;9) means AML with translocation (t6;9)

der(5)t(5;11) means AML with translocation (t5;11)

i(17) means AML with isochromosome 17

del(9q) means AML with deletion on Chromosome 9q

del(12p)) means AML with deletion on chromosome 12p

del(20q) means AML with deletion on Chromosome 20q.

As used herein, “all other subtypes” refer to the subtypes of the present invention, i.e. if one subtype is distinguished from “all other subtypes”, it is distiguished from all other subtypes contained in the present invention.

According to the present invention, a “sample” means any biological material containing genetic information in the form of nucleic acids or proteins obtainable or obtained from an individual. The sample includes e.g. tissue samples, cell samples, bone marrow and/or body fluids such as blood, saliva, semen. Preferably, the sample is blood or bone marrow, more preferably the sample is bone marrow. The person skilled in the art is aware of methods, how to isolate nucleic acids and proteins from a sample. A general method for isolating and preparing nucleic acids from a sample is outlined in Example 3.

According to the present invention, the term “lower expression” is generally assigned to all by numbers and Affymetrix Id. definable polynucleotides the t-values and fold change (fc) values of which are negative, as indicated in the Tables. Accordingly, the term “higher expression” is generally assigned to all by numbers and Affymetrix Id. definable polynucleotides the t-values and fold change (fc) values of which are positive.

According to the present invention, the term “lower expression” is generally assigned to all by numbers and Affymetrix Id. definable polynucleotides the t-values and fold change (fc) values of which are negative, as indicated in the Tables. Accordingly, the term “higher expression” is generally assigned to all by numbers and Affymetrix Id. definable polynucleotides the t-values and fold change (fc) values of which are positive.

According to the present invention, the term "expression" refers to the process by which mRNA or a polypeptide is produced based on the nucleic acid sequence of a gene, i.e. „expression“ also includes the formation of mRNA upon transcription. In accordance with the present invention, the term „determining the expression level” preferably refers to the determination of the level of expression, namely of the markers.

Generally, “marker” refers to any genetically controlled difference which can be used in the genetic analysis of a test versus a control sample, for the purpose of assigning the sample to a defined genotype or phenotype. As used herein, “markers” refer to genes

which are differentially expressed in, e.g., different AML subtypes. The markers can be defined by their gene symbol name, their encoded protein name, their transcript identification number (cluster identification number), the data base accession number, public accession number or GenBank identifier or, as done in the present invention, Affymetrix identification number, chromosomal location, UniGene accession number and cluster type, LocusLink accession number (see Examples and Tables).

The Affymetrix identification number (affy id) is accessible for anyone and the person skilled in the art by entering the "gene expression omnibus" internet page of the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/geo/). In particular, the affy id's of the polynucleotides used for the method of the present invention are derived from the so-called U133 chip. The sequence data of each identification number can be viewed at http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL96

Generally, the expression level of a marker is determined by the determining the expression of its corresponding "polynucleotide" as described hereinafter.

According to the present invention, the term „polynucleotide" refers, generally, to a DNA, in particular cDNA, or RNA, in particular a cRNA, or a portion thereof or a polypeptide or a portion thereof. In the case of RNA (or cDNA), the polynucleotide is formed upon transcription of a nucleotide sequence which is capable of expression. The polynucleotide fragments refer to fragments preferably of between at least 8, such as 10, 12, 15 or 18 nucleotides and at least 50, such as 60, 80, 100, 200 or 300 nucleotides in length, or a complementary sequence thereto, representing a consecutive stretch of nucleotides of a gene, cDNA or mRNA. In other terms, polynucleotides include also any fragment (or complementary sequence thereto) of a sequence derived from any of the markers defined above as long as these fragments unambiguously identify the marker.

The determination of the expression level may be effected at the transcriptional or translational level, i.e. at the level of mRNA or at the protein level. Protein fragments such as peptides or polypeptides advantageously comprise between at least 6 and at least 25, such as 30, 40, 80, 100 or 200 consecutive amino acids representative of the corresponding full length protein. Six amino acids are generally recognized as the lowest peptidic stretch giving rise to a linear epitope recognized by an antibody, fragment or derivative thereof.

Alternatively, the proteins or fragments thereof may be analysed using nucleic acid molecules specifically binding to three-dimensional structures (aptamers).

Depending on the nature of the polynucleotide or polypeptide, the determination of the expression levels may be effected by a variety of methods. For determining and detecting the expression level, it is preferred in the present invention that the polynucleotide, in particular the cRNA, is labelled.

The labelling of the polynucleotide or a polypeptide can occur by a variety of methods known to the skilled artisan. The label can be fluorescent, chemiluminescent, bioluminescent, radioactive (such as ^{3}H or ^{32}P). The labelling compound can be any labelling compound being suitable for the labelling of polynucleotides and/or polypeptides. Examples include fluorescent dyes, such as fluorescein, dichlorofluorescein, hexachlorofluorescein, BODIPY variants, ROX, tetramethylrhodamin, rhodamin X, Cyanine-2, Cyanine-3, Cyanine-5, Cyanine-7, IRD40, FluorX, Oregon Green, Alexa variants (available e.g. from Molecular Probes or Amersham Biosciences) and the like, biotin or biotinylated nucleotides, digoxigenin, radioisotopes, antibodies, enzymes and receptors. Depending on the type of labelling, the detection is done via fluorescence measurements, conjugation to streptavidin and/or avidin, antigen-antibody- and/or antibody-antibody-interactions, radioactivity measurements, as well as catalytic and/or receptor/ligand interactions. Suitable methods include the direct labelling (incorporation) method, the amino-modified (amino-allyl) nucleotide method (available e.g. from Ambion), and the primer tagging method (DNA dendrimer labelling, as kit available e.g. from Genisphere). Particularly preferred for the present invention is the use of biotin or biotinylated nucleotides for labelling, with the latter being directly incorporated into, e.g. the cRNA polynucleotide by in vitro transcription.

If the polynucleotide is mRNA, cDNA may be prepared into which a detectable label, as exemplified above, is incorporated. Said detectably labelled cDNA, in single-stranded form, may then be hybridised, preferably under stringent or highly stringent conditions to a panel of single-stranded oligonucleotides representing different genes and affixed to a solid support such as a chip. Upon applying appropriate washing steps, those cDNAs will be detected or quantitatively detected that have a counterpart in the oligonucleotide panel. Various advantageous embodiments of this general method are feasible. For example, the mRNA or the cDNA may be amplified e.g. by polymerase chain reaction, wherein it is preferable, for quantitative assessments, that the number of amplified copies corresponds

relative to further amplified mRNAs or cDNAs to the number of mRNAs originally present in the cell. In a preferred embodiment of the present in ivention, the cDNAs are transcribed into cRNAs prior to the hybridisation step wherein only in the transcription step a label is incorporated into the nucleic acid and wherein the cRNA is employed for hybridisation. Alternatively, the label may be attached subsequent to the transcription step.

Similarly, proteins from a cell or tissue under investigation may be contacted with a panel of aptamers or of antibodies or fragments or derivatives thereof. The antibodies etc. may be affixed to a solid support such as a chip. Binding of proteins indicative of an AML subtype may be verified by binding to a detectably labelled secondary antibody or aptamer. For the labelling of antibodies, it is referred to Harlow and Lane, "Antibodies, a laboratory manual", CSH Press, 1988, Cold Spring Harbor. Specifically, a minimum set of proteins necessary for diagnosis of all AML subtypes may be selected for creation of a protein array system to make diagnosis on a protein lysate of a diagnostic bone marrow sample directly. Protein Array Systems for the detection of specific protein expression profiles already are available (for example: Bio-Plex, BIORAD, München, Germany). For this application preferably antibodies against the proteins have to be produced and immobilized on a platform e.g. glasslides or microtiterplates. The immobilized antibodies can be labelled with a reactant specific for the certain target proteins as discussed above. The reactants can include enzyme substrates, DNA, receptors, antigens or antibodies to create for example a capture sandwich immunoassay.

For reliably distinguishing AML subtypes with aberrant and prognostically intermediate karyotypes with different prognosis it is useful that the expression of more than one of the above defined markers is determined. As a criterion for the choice of markers, the statistical significance of markers as expressed in q or p values based on the concept of the false discovery rate is determined. In doing so, a measure of statistical significance called the *q* value is associated with each tested feature. The *q* value is similar to the *p* value, except it is a measure of significance in terms of the false discovery rate rather than the false positive rate (Storey JD and Tibshirani R. Proc.Natl.Acad.Sci., 2003, Vol. 100:9440-5.

In a preferred embodiment of the present invention, markers as defined in Table 1-2 having a q-value of less than 3E-06, more preferred less than 1.5E-09, most preferred less than 1.5E-11, less than 1.5E-20, less than 1.5E-30, are measured.

Of the above defined markers, the expression level of at least two, preferably of at least ten, more preferably of at least 25, most preferably of 50 of at least one of the Tables of the markers is determined.

In another preferred embodiment, the expression level of at least 2, of at least 5, of at least 10 out of the markers having the numbers 1 – 10, 1-20, 1-40, 1-50 of at least one of the Tables are measured.

The level of the expression of the „marker", i.e. the expression of the polynucleotide is indicative of the AML subtype of a cell or an organism. The level of expression of a marker or group of markers is measured and is compared with the level of expression of the same marker or the same group of markers from other cells or samples. The comparison may be effected in an actual experiment or in silico. When the expression level also referred to as expression pattern or expression signature (expression profile) is measurably different, there is according to the invention a meaningful difference in the level of expression. Preferably the difference at least is 5 %, 10% or 20%, more preferred at least 50% or may even be as high as 75% or 100%. More preferred the difference in the level of expression is at least 200%, i.e. two fold, at least 500%, i.e. five fold, or at least 1000%, i.e. 10 fold.

Accordingly, the expression level of markers expressed lower in a first subtype than in at least one second subtype, which differs from the first subtype, is at least 5 %, 10% or 20%, more preferred at least 50% or may even be 75% or 100%, i.e. 2-fold lower, preferably at least 10-fold, more preferably at least 50-fold, and most preferably at least 100-fold lower in the first subtype. On the other hand, the expression level of markers expressed higher in a first subtype than in at least one second subtype, which differs from the first subtype, is at least 5 %, 10% or 20%, more preferred at least 50% or may even be 75% or 100%, i.e. 2-fold higher, preferably at least 10-fold, more preferably at least 50-fold, and most preferably at least 100-fold higher in the first subtype.

In another embodiment of the present invention, the sample is derived from an individual having leukaemia, preferably AML.

For the method of the present invention it is preferred if the polynucleotide the expression level of which is determined is in form of a transcribed polynucleotide. A particularly

preferred transcribed polynucleotide is an mRNA, a cDNA and/or a cRNA, with the latter being preferred. Transcribed polynucleotides are isolated from a sample, reverse transcribed and/or amplified, and labelled, by employing methods well-known the person skilled in the art (see Example 3). In a preferred embodiment of the methods according to the invention, the step of determining the expression profile further comprises amplifying the transcribed polynucleotide.

In order to determine the expression level of the transcribed polynucleotide by the method of the present invention, it is preferred that the method comprises hybridizing the transcribed polynucleotide to a complementary polynucleotide, or a portion thereof, under stringent hybridization conditions, as described hereinafter.

The term "hybridizing" means hybridization under conventional hybridization conditions, preferably under stringent conditions as described, for example, in Sambrook, J., et al., in "Molecular Cloning: A Laboratory Manual" (1989), Eds. J. Sambrook, E. F. Fritsch and T. Maniatis, Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY and the further definitions provided above. Such conditions are, for example, hybridization in 6x SSC, pH 7.0 / 0.1% SDS at about 45°C for 18-23 hours, followed by a washing step with 2x SSC/0.1% SDS at 50°C. In order to select the stringency, the salt concentration in the washing step can for example be chosen between 2x SSC/0.1% SDS at room temperature for low stringency and 0.2x SSC/0.1% SDS at 50°C for high stringency. In addition, the temperature of the washing step can be varied between room temperature, ca. 22°C, for low stringency, and 65°C to 70° C for high stringency. Also contemplated are polynucleotides that hybridize at lower stringency hybridization conditions. Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation, preferably of formamide concentration (lower percentages of formamide result in lowered stringency), salt conditions, or temperature. For example, lower stringency conditions include an overnight incubation at 37°C in a solution comprising 6X SSPE (20X SSPE = 3M NaCl; 0.2M NaH2PO4; 0.02M EDTA, pH 7.4), 0.5% SDS, 30% formamide, 100 mg/ml salmon sperm blocking DNA, followed by washes at 50°C with 1 X SSPE, 0.1% SDS. In addition, to achieve even lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5x SSC). Variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility.

"Complementary" and "complementarity", respectively, can be described by the percentage, i.e. proportion, of nucleotides which can form base pairs between two polynucleotide strands or within a specific region or domain of the two strands. Generally, complementary nucleotides are, according to the base pairing rules, adenine and thymine (or adenine and uracil), and cytosine and guanine. Complementarity may be partial, in which only some of the nucleic acids' bases are matched according to the base pairing rules. Or, there may be a complete or total complementarity between the nucleic acids. The degree of complementarity between nucleic acid strands has effects on the efficiency and strength of hybridization between nucleic acid strands.

Two nucleic acid strands are considered to be 100% complementary to each other over a defined length if in a defined region all adenines of a first strand can pair with a thymine (or an uracil) of a second strand, all guanines of a first strand can pair with a cytosine of a second strand, all thymine (or uracils) of a first strand can pair with an adenine of a second strand, and all cytosines of a first strand can pair with a guanine of a second strand, and vice versa. According to the present invention, the degree of complementarity is determined over a stretch of 20, preferably 25, nucleotides, i.e. a 60% complementarity means that within a region of 20 nucleotides of two nucleic acid strands 12 nucleotides of the first strand can base pair with 12 nucleotides of the second strand according to the above ruling, either as a stretch of 12 contiguous nucleotides or interspersed by non-pairing nucleotides, when the two strands are attached to each other over said region of 20 nucleotides. The degree of complementarity can range from at least about 50% to full, i.e. 100% complementarity. Two single nucleic acid strands are said to be "substantially complementary" when they are at least about 80% complementary, preferably about 90% or higher. For carrying out the method of the present invention substantial complementarity is preferred.

Preferred methods for detection and quantification of the amount of polynucleotides, i.e. for the methods according to the invention allowing the determination of the level of expression of a marker, are those described by Sambrook et al. (1989) or real time methods known in the art as the TaqMan® method disclosed in WO92/02638 and the corresponding U.S. 5,210,015, U.S. 5,804,375, U.S. 5,487,972. This method exploits the exonuclease activity of a polymerase to generate a signal. In detail, the (at least one) target nucleic acid component is detected by a process comprising contacting the sample with an oligonucleotide containing a sequence complementary to a region of the target nucleic acid component and a labeled oligonucleotide containing a sequence complementary to a

second region of the same target nucleic acid component sequence strand, but not including the nucleic acid sequence defined by the first oligonucleotide, to create a mixture of duplexes during hybridization conditions, wherein the duplexes comprise the target nucleic acid annealed to the first oligonucleotide and to the labeled oligonucleotide such that the 3'-end of the first oligonucleotide is adjacent to the 5'-end of the labeled oligonucleotide. Then this mixture is treated with a template-dependent nucleic acid polymerase having a 5' to 3' nuclease activity under conditions sufficient to permit the 5' to 3' nuclease activity of the polymerase to cleave the annealed, labeled oligonucleotide and release labeled fragments. The signal generated by the hydrolysis of the labeled oligonucleotide is detected and/ or measured. TaqMan® technology eliminates the need for a solid phase bound reaction complex to be formed and made detectable. Other methods include e.g. fluorescence resonance energy transfer between two adjacently hybridized probes as used in the LightCycler® format described in U.S. 6,174,670.

A preferred protocol if the marker, i.e. the polynucleotide, is in form of a transcribed nucleotide, is described in Example 3, where total RNA is isolated, cDNA and, subsequently, cRNA is synthesized and biotin is incorporated during the transcription reaction. The purified cRNA is applied to commercially available arrays which can be obtained e.g. from Affymetrix. The hybridized cRNA is detected according to the methods described in Example 3. The arrays are produced by photolithography or other methods known to experts skilled in the art e.g. from U.S. 5,445,934, U.S. 5,744,305, U.S. 5,700,637, U.S. 5,945,334 and EP 0 619 321 or EP 0 373 203, or as decribed hereinafter in greater detail.

In another embodiment of the present invention, the polynucleotide or at least one of the polynucleotides is in form of a polypeptide. In another preferred embodiment, the expression level of the polynucleotides or polypeptides is detected using a compound which specifically binds to the polynucleotide of the polypeptide of the present invention.

As used herein, "specifically binding" means that the compound is capable of discriminating between two or more polynucleotides or polypeptides, i.e. it binds to the desired polynucleotide or polypeptide, but essentially does not bind unspecifically to a different polynucleotide or polypeptide.

The compound can be an antibody, or a fragment thereof, an enzyme, a so-called small molecule compound, a protein-scaffold, preferably an anticalin. In a preferred

embodiment, the compound specifically binding to the polynucleotide or polypeptide is an antibody, or a fragment thereof.

As used herein, an "antibody" comprises monoclonal antibodies as first described by Köhler and Milstein in Nature 278 (1975), 495-497 as well as polyclonal antibodies, i.e. entibodies contained in a polyclonal antiserum. Monoclonal antibodies include those produced by transgenic mice. Fragments of antibodies include $F(ab')_2$, Fab and Fv fragments. Derivatives of antibodies include scFvs, chimeric and humanized antibodies. See, for example Harlow and Lane, loc. cit. For the detection of polypeptides using antibodies or fragments thereof, the person skilled in the art is aware of a variety of methods, all of which are included in the present invention. Examples include immunoprecipitation, Western blotting, Enzyme-linked immuno sorbent assay (ELISA), Enzyme-linked immuno sorbent assay (RIA), dissociation-enhanced lanthanide fluoro immuno assay (DELFIA), scintillation proximity assay (SPA). For detection, it is desirable if the antibody is labelled by one of the labelling compounds and methods described supra.

In another preferred embodiment of the present invention, the method for distinguishing AML subtypes with aberrant and prognostically intermediate karyotypes into different prognostic subsets is carried out on an array.

In general, an "array" or "microarray" refers to a linear or two- or three dimensional arrangement of preferably discrete nucleic acid or polypeptide probes which comprises an intentionally created collection of nucleic acid or polypeptide probes of any length spotted onto a substrate/solid support. The person skilled in the art knows a collection of nucleic acids or polypeptide spotted onto a substrate/solid support also under the term "array". As known to the person skilled in the art, a microarray usually refers to a miniaturised array arrangement, with the probes being attached to a density of at least about 10, 20, 50, 100 nucleic acid molecules referring to different or the same genes per cm^2. Furthermore, where appropriate an array can be referred to as "gene chip". The array itself can have different formats, e.g. libraries of soluble probes or libraries of probes tethered to resin beads, silica chips, or other solid supports.

The process of array fabrication is well-known to the person skilled in the art. In the following, the process for preparing a nucleic acid array is described. Commonly, the process comprises preparing a glass (or other) slide (e.g. chemical treatment of the glass to

enhance binding of the nucleic acid probes to the glass surface), obtaining DNA sequences representing genes of a genome of interest, and spotting sequences these sequences of interest onto glass slide. Sequences of interest can be obtained via creating a cDNA library from an mRNA source or by using publicly available databases, such as GeneBank, to annotate the sequence information of custom cDNA libraries or to identify cDNA clones from previously prepared libraries. Generally, it is recommendable to amplify obtained sequences by PCR in order to have sufficient amounts of DNA to print on the array. The liquid containing the amplified probes can be deposited on the array by using a set of microspotting pins. Ideally, the amount deposited should be uniform. The process can further include UV-crosslinking in order to enhance immobilization of the probes on the array.

In a preferred embodiment, the array is a high density oligonucleotide (oligo) array using a light-directed chemical synthesis process, employing the so-called photolithography technology. Unlike common cDNA arrays, oligo arrays (according to the Affymetrix technology) use a single-dye technology. Given the sequence information of the markers, the sequence can be synthesized directly onto the array, thus, bypassing the need for physical intermediates, such as PCR products, required for making cDNA arrays. For this purpose, the marker, or partial sequences thereof, can be represented by 14 to 20 features, preferably by less than 14 features, more preferably less than 10 features, even more preferably by 6 features or less, with each feature being a short sequence of nucleotides (oligonucleotide), which is a perfect match (PM) to a segment of the respective gene. The PM oligonucleotide are paired with mismatch (MM) oligonucleotides which have a single mismatch at the central base of the nucleotide and are used as "controls". The chip exposure sites are defined by masks and are deprotected by the use of light, followed by a chemical coupling step resulting in the synthesis of one nucleotide. The masking, light deprotection, and coupling process can then be repeated to synthesize the next nucleotide, until the nucleotide chain is of the specified length.

Advantageously, the method of the present invention is carried out in a robotics system including robotic plating and a robotic liquid transfer system, e.g. using microfluidics, i.e. channelled structured.

A particular preferred method according to the present invention is as follows:

1. Obtaining a sample, e.g. bone marrow or peripheral blood aliquots, from a patient having AML
2. Extracting RNA, preferably mRNA, from the sample

3. Reverse transcribing the RNA into cDNA
4. In vitro transcribing the cDNA into cRNA
5. Fragmenting the cRNA
6. Hybridizing the fragmented cRNA on standard microarrays
7. Determining hybridization

In another embodiment, the present invention is directed to the use of at least one marker selected from the markers identifiable by their Affymetrix Identification Numbers (affy id) as defined in Tables 1, and/or 2 for the manufacturing of a diagnostic for distinguishing AML subtypes with aberrant and prognostically intermediate karyotypes. The use of the present invention is particularly advantageous for distinguishing AML subtypes with aberrant and prognostically intermediate karyotypes in an individual having AML. The use of said markers for diagnosis of AML subtypes with aberrant and prognostically intermediate karyotypes, preferably based on microarray technology, offers the following advantages: (1) more rapid and more precise diagnosis, (2) easy to use in laboratories without specialized experience, (3) abolishes the requirement for analyzing viable cells for chromosome analysis (transport problem), and (4) very experienced hematologists for cytomorphology and cytochemistry, immunophenotyping as well as cytogeneticists and molecularbiologists are no longer required.

Accordingly, the present invention refers to a diagnostic kit containing at least one marker selected from the markers identifiable by their Affymetrix Identification Numbers (affy id) as defined in Tables 1, and/or 2 for distinguishing AML subtypes with aberrant and prognostically intermediate karyotypes, in combination with suitable auxiliaries. Suitable auxiliaries, as used herein, include buffers, enzymes, labelling compounds, and the like. In a preferred embodiment, the marker contained in the kit is a nucleic acid molecule which is capable of hybridizing to the mRNA corresponding to at least one marker of the present invention. Preferably, the at least one nucleic acid molecule is attached to a solid support, e.g. a polystyrene microtiter dish, nitrocellulose membrane, glass surface or to non-immobilized particles in solution.

In another preferred embodiment, the diagnostic kit contains at least one reference for an AML subtype with aberrant and prognostically intermediate karyotypes. As used herein, the reference can be a sample or a data bank.

In another embodiment, the present invention is directed to an apparatus for distinguishing AML subtypes with aberrant and prognostically intermediate karyotypes selected from

trisomy 8, inv(3), t(3;3), trisomy 11, trisomy 13, trisomy 4, t(1;3), t(6;9), der(5)t(5;11), i(17), del(9q), del(12p), and/or del(20q) in a sample, containing a reference data bank obtainable by comprising

(a) compiling a gene expression profile of a patient sample by determining the expression level at least one marker selected from the markers identifiable by their Affymetrix Identification Numbers (affy id) as defined in Tables 1, and/or 2, and

(b) classifying the gene expression profile by means of a machine learning algorithm.

According to the present invention, the "machine learning algorithm" is a computational-based prediction methodology, also known to the person skilled in the art as "classifier", employed for characterizing a gene expression profile. The signals corresponding to a certain expression level which are obtained by the microarray hybridization are subjected to the algorithm in order to classify the expression profile. Supervised learning involves "training" a classifier to recognize the distinctions among classes and then "testing" the accuracy of the classifier on an independent test set. For new, unknown sample the classifier shall predict into which class the sample belongs.

Preferably, the machine learning algorithm is selected from the group consisting of Weighted Voting, K-Nearest Neighbors, Decision Tree Induction, Support Vector Machines (SVM), and Feed-Forward Neural Networks. Most preferably, the machine learning algorithm is Support Vector Machine, such as polynomial kernel and Gaussian Radial Basis Function-kernel SVM models.

The classification accuracy of a given gene list for a set of microarray experiments is preferably estimated using Support Vector Machines (SVM), because there is evidence that SVM-based prediction slightly outperforms other classification techniques like k-Nearest Neighbors (k-NN). The LIBSVM software package version 2.36 was used (SVM-type: C-SVC, linear kernel (http://www.csie.ntu.edu.tw/~cjlin/libsvm/)). The skilled artisan is furthermore referred to Brown et al., Proc.Natl.Acad.Sci., 2000; 97: 262-267, Furey et al., Bioinformatics. 2000; 16: 906-914, and Vapnik V. Statistical Learning Theory. New York: Wiley, 1998.

In detail, the classification accuracy of a given gene list for a set of microarray experiments can be estimated using Support Vector Machines (SVM) as supervised learning technique.

Generally, SVMs are trained using differentially expressed genes which were identified on a subset of the data and then this trained model is employed to assign new samples to those trained groups from a second and different data set. Differentially expressed genes were identified applying ANOVA and t-test-statistics (Welch t-test). Based on identified distinct gene expression signatures respective training sets consisting of 2/3 of cases and test sets with 1/3 of cases to assess classification accuracies are designated. Assignment of cases to training and test set is randomized and balanced by diagnosis. Based on the training set a Support Vector Machine (SVM) model is built.

According to the present invention, the apparent accuracy, i.e. the overall rate of correct predictions of the complete data set was estimated by 10fold cross validation. This means that the data set was divided into 10 approximately equally sized subsets, an SVM-model was trained for 9 subsets and predictions were generated for the remaining subset. This training and prediction process was repeated 10 times to include predictions for each subset. Subsequently the data set was split into a training set, consisting of two thirds of the samples, and a test set with the remaining one third. Apparent accuracy for the training set was estimated by 10fold cross validation (analogous to apparent accuracy for complete set). A SVM-model of the training set was built to predict diagnosis in the independent test set, thereby estimating true accuracy of the prediction model. This prediction approach was applied both for overall classification (multi-class) and binary classification (diagnosis X => yes or no). For the latter, sensitivity and specificity were calculated:

Sensitivity = (number of positive samples predicted)/(number of true positives)

Specificity = (number of negative samples predicted)/(number of true negatives)

In a preferred embodiment, the reference data bank is backed up on a computational data memory chip which can be inserted in as well as removed from the apparatus of the present invention, e.g. like an interchangeable module, in order to use another data memory chip containing a different reference data bank.

The apparatus of the present invention containing a desired reference data bank can be used in a way such that an unknown sample is, first, subjected to gene expression profiling, e.g. by microarray analysis in a manner as described supra or in the art, and the expression level data obtained by the analysis are, second, fed into the apparatus and compared with

the data of the reference data bank obtainable by the above method. For this purpose, the apparatus suitably contains a device for entering the expression level of the data, for example a control panel such as a keyboard. The results, whether and how the data of the unknown sample fit into the reference data bank can be made visible on a provided monitor or display screen and, if desired, printed out on an incorporated of connected printer.

Alternatively, the apparatus of the present invention is equipped with particular appliances suitable for detecting and measuring the expression profile data and, subsequently, proceeding with the comparison with the reference data bank. In this embodiment, the apparatus of the present invention can contain a gripper arm and/or a tray which takes up the microarray containing the hybridized nucleic acids.

In another embodiment, the present invention refers to a reference data bank for distinguishing AML subtypes with aberrant and prognostically intermediate karyotypes selected from trisomy 8, inv(3), t(3;3), trisomy 11, trisomy 13, trisomy 4, t(1;3), t(6;9), der(5)t(5;11), i(17), del(9q), del(12p), and/or del(20q) in a sample obtainable by comprising

(a) compiling a gene expression profile of a patient sample by determining the expression level of at least one marker selected from the markers identifiable by their Affymetrix Identification Numbers (affy id) as defined in Tables 1, and/or 2, and
(b) classifying the gene expression profile by means of a machine learning algorithm.

Preferably, the reference data bank is backed up and/or contained in a computational memory data chip.

The invention is further illustrated in the following table and examples, without limiting the scope of the invention:

TABLES 1-2

Tables 1-2 show AML subtype analysis of AML subtypes with aberrant and prognostically intermediate karyotypes selected from trisomy 8, inv(3), t(3;3), trisomy 11, trisomy 13, trisomy 4, t(1;3), t(6;9), der(5)t(5;11), i(17), del(9q), del(12p), and/or del(20q). The

analysed markers are ordered according to their q- and p-values, beginning with the lowest q and p-values.

zres is the test statistic of the Cox model. It is calculated as the ratio of the estimated cox regression coefficient and its estimated standard error.

P. Andersen and R. Gill: "Cox's regression model for counting processes, a large sample study", Annals of Statistics, 10:1100-1120, 1982.

Pres is non-corrected p-value
Qres is corrected p-value

For convenience and a better understanding, Tables 1-2 are accompanied with explanatory tables (Table 1A-2A) where the numbering and the Affymetrix Id are further defined by other parameters, e.g. gene bank accession number.

EXAMPLES

Example 1: General experimental design of the invention and results

The karyotype defines the biology of distinct subtypes in acute myeloid leukaemia (AML) and is the most important prognostic factor. However, the cytogenetically based classification of AML reveals a large group with prognostically intermediate or undetermined karyotypes. Within these cases those without chromosomal aberrations may be further characterized by molecular genetic alterations like length mutations of the FLT3 gene or partial tandem duplication of the MLL gene. These mutations are present in half of these patients and have been proven prognostically relevant. In contrast, for patients with aberrant and prognostically intermediate and undetermined karyotype no genetic markers have been capable of predicting the patients outcome. Using Affymetrix U113A+B microarrays we determined the expression status of more than 30,000 genes in 25 patients with AML and karyotype aberrations rated as prognostically intermediate and undetermined (median age 55 years, range 18-77). Karyotype abnormalities included trisomy 8 as sole abnormality (n=7), inv(3) or t(3;3) (n=6), trisomy 11 (n=2), trisomy 13 (n=2), and trisomy 4, t(1;3), t(6;9), der(5)t(5;11), i(17), del(9q), del(12p), del(20q) (n=1 for each aberration). All patients were uniformly treated within the 1999 trial of the German AMLCG (TAD/HAM or HAM/HAM double induction therapy, TAD consolidation

therapy, maintenance therapy). Median event-free survival (EFS) was 3.1 months and median overall survival (OS) was 20.7 months. The clinical parameters age, WBC count, hemoglobin level, and thrombocyte count were not related to either EFS or OS. Univariate Cox regression analyses were performed to identify genes significantly related to EFS and OS. Among the top 50 genes/transcripts related to EFS five were identified by multivariate analysis to independenty influence EFS: PRAME (p=0.037); human BRCA2 region, mRNA sequence CG006 (p=0.042); Homo sapiens cDNA: FLJ22765 fis, clone KAIA1180 (p=0.031); Homo sapiens cDNA FLJ36837 fis, clone ASTRO2011422 (p=0.032); and ESTs (p=0.049). A score was defined based on the expression status of these genes. This score revealed three groups (0 out of 5 vs. 1 to 3 out of 5 vs. 4 to 5 out of 5 genes expressed) with significantly differing EFS (75% at 1 year vs. median 3.1 months vs. median 1.2 months, p<0.0001). This score also resulted in significant differences in OS (100% at 1 year vs. median 20.7 months vs. median 1.7 months, p=0.0148). To check for the consistency of these prognostic genes within AML with aberrant and prognostically intermediate karyotypes 25 additional cases with other aberrations were analyzed (monosomy 7 n=6, t(3;21) n=1, t(6;9) n=1, del(5q) n=2, del(9q) n=2, inv(3) and monosomy 7 n=2, t(3;3) and t(5;17) n=1, trisomy 11 n=3, trisomy 13 n=5, trisomy 8 n=2). However, the score had no prognostic impact in this cohort. These data demonstrate that prognostically relevant genes may be identified in AML cases in which at present no prognostic markers are available. It is suggested that a combination of the expression status of multiple genes is necessary to accomplish prognostication which is in line with the concept of a multi-genetic basis of the leukemogenesis in these cases. Furthermore, it is suggested that for optimizing the performance of genetically based prognostic scores these should be applied only to cytogenetically homogeneous cohorts.

Example 2: General materials, methods and definitions of functional annotations

The methods section contains both information on statistical analyses used for identification of differentially expressed genes and detailed annotation data of identified microarray probesets.

Affymetrix Probeset Annotation

All annotation data of GeneChip® arrays are extracted from the NetAffx™ Analysis Center (internet website: www.affymetrix.com). Files for U133 set arrays, including U133A and U133B microarrays are derived from the June 2003 release. The original

publication refers to: Liu G, Loraine AE, Shigeta R, Cline M, Cheng J, Valmeekam V, Sun S, Kulp D, Siani-Rose MA. NetAffx: Affymetrix probesets and annotations. Nucleic Acids Res. 2003;31(1):82-6.

The sequence data are omitted due to their large size, and because they do not change, whereas the annotation data are updated periodically, for example new information on chromomal location and functional annotation of the respective gene products. Sequence data are available for download in the NetAffx Download Center (www.affymetrix.com)

Data fields:
In the following section, the content of each field of the data files are described. Microarray probesets, for example found to be differentially expressed between different types of leukemia samples are further described by additional information. The fields are of the following types:

1. GeneChip Array Information
2. Probe Design Information
3. Public Domain and Genomic References

1. GeneChip Array Information

HG-U133 ProbeSet_ID:
HG-U133 ProbeSet_ID describes the probe set identifier. Examples are: 200007_at, 200011_s_at, 200012_x_at.

GeneChip:
The description of the GeneChip probe array name where the respective probeset is represented. Examples are: Affymetrix Human Genome U133A Array or Affymetrix Human Genome U133B Array.

2. Probe Design Information

Sequence Type:
The Sequence Type indicates whether the sequence is an Exemplar, Consensus or Control sequence. An Exemplar is a single nucleotide sequence taken directly from a public

database. This sequence could be an mRNA or EST. A Consensus sequence, is a nucleotide sequence assembled by Affymetrix, based on one or more sequence taken from a public database.

Transcript ID:
The cluster identification number with a sub-cluster identifier appended.

Sequence Derived From:
The accession number of the single sequence, or representative sequence on which the probe set is based. Refer to the "Sequence Source" field to determine the database used.

Sequence ID:
For Exemplar sequences: Public accession number or GenBank identifier. For Consensus sequences: Affymetrix identification number or public accession number.

Sequence Source:
The database from which the sequence used to design this probe set was taken. Examples are: GenBank®, RefSeq, UniGene, TIGR (annotations from The Institute for Genomic Research).

3. Public Domain and Genomic References

Most of the data in this section come from LocusLink and UniGene databases, and are annotations of the reference sequence on which the probe set is modeled.

Gene Symbol and Title:
A gene symbol and a short title, when one is available. Such symbols are assigned by different organizations for different species. Affymetrix annotational data come from the UniGene record. There is no indication which species-specific databank was used, but some of the possibilities include for example HUGO: The Human Genome Organization.

MapLocation:
The map location describes the chromosomal location when one is available.

Unigene_Accession:
UniGene accession number and cluster type. Cluster type can be "full length" or "est", or "---" if unknown.

LocusLink:
This information represents the LocusLink accession number.

Full Length Ref. Sequences:
Indicates the references to multiple sequences in RefSeq. The field contains the ID and description for each entry, and there can be multiple entries per probeSet.

Example 3: Sample preparation, processing and data analysis

Method 1:

Microarray analyses were performed utilizing the GeneChip® System (Affymetrix, Santa Clara, USA). Hybridization target preparations were performed according to recommended protocols (Affymetrix Technical Manual). In detail, at time of diagnosis, mononuclear cells were purified by Ficoll-Hypaque density centrifugation. They had been lysed immediately in RLT buffer (Qiagen, Hilden, Germany), frozen, and stored at -80°C from 1 week to 38 months. For gene expression profiling cell lysates of the leukemia samples were thawed, homogenized (QIAshredder, Qiagen), and total RNA was extracted (RNeasy Mini Kit, Qiagen). Subsequently, 5-10 µg total RNA isolated from 1 x 10^7 cells was used as starting material for cDNA synthesis with oligo[$(dT)_{24}$T7promotor]$_{65}$ primer (cDNA Synthesis System, Roche Applied Science, Mannheim, Germany). cDNA products were purified by phenol/chlorophorm/IAA extraction (Ambion, Austin, USA) and acetate/ethanol-precipitated overnight. For detection of the hybridized target nucleic acid biotin-labeled ribonucleotides were incorporated during the following *in vitro* transcription reaction (Enzo BioArray HighYield RNA Transcript Labeling Kit, Enzo Diagnostics). After quantification by spectrophotometric measurements and 260/280 absorbance values assessment for quality control of the purified cRNA (RNeasy Mini Kit, Qiagen), 15 µg cRNA was fragmented by alkaline treatment (200 mM Tris-acetate, pH 8.2/500 mM potassium acetate/150 mM magnesium acetate) and added to the hybridization cocktail sufficient for five hybridizations on standard GeneChip microarrays (300 µl final volume). Washing and staining of the probe arrays was performed according to the recommended Fluidics Station protocol (EukGE-WS2v4). Affymetrix Microarray Suite software (version

5.0.1) extracted fluorescence signal intensities from each feature on the microarrays as detected by confocal laser scanning according to the manufacturer's recommendations.

Expression analysis quality assessment parameters included visiual array inspection of the scanned image for the presence of image artifacts and correct grid alignment for the identification of distinct probe cells as well as both low 3'/5' ratio of housekeeping controls (mean: 1.90 for GAPDH) and high percentage of detection calls (mean: 46.3% present called genes). The 3' to 5' ratio of GAPDH probesets can be used to assess RNA sample and assay quality. Signal values of the 3' probe sets for GAPDH are compared to the Signal values of the corresponding 5' probe set. The ratio of the 3' probe set to the 5' probe set is generally no more than 3.0. A high 3' to 5' ratio may indicate degraded RNA or inefficient synthesis of ds cDNA or biotinylated cRNA (GeneChip® Expression Analysis Technical Manual, www.affymetrix.com). Detection calls are used to determine whether the transcript of a gene is detected (present) or undetected (absent) and were calculated using default parameters of the Microarray Analysis Suite MAS 5.0 software package.

Method 2:

Bone marrow (BM) aspirates are taken at the time of the initial diagnostic biopsy and remaining material is immediately lysed in RLT buffer (Qiagen), frozen and stored at -80 C until preparation for gene expression analysis. For microarray analysis the GeneChip System (Affymetrix, Santa Clara, CA, USA) is used. The targets for GeneChip analysis are prepared according to the current Expression Analysis. Briefly, frozen lysates of the leukemia samples are thawed, homogenized (QIAshredder, Qiagen) and total RNA extracted (RNeasy Mini Kit, Qiagen).Normally 10 ug total RNA isolated from 1 x 107 cells is used as starting material in the subsequent cDNA-Synthesis using Oligo-dT-T7-Promotor Primer (cDNA synthesis Kit, Roche Molecular Biochemicals). The cDNA is purified by phenol-chlorophorm extraction and precipitated with 100% Ethanol over night. For detection of the hybridized target nucleic acid biotin-labeled ribonucleotides are incorporated during the in vitro transcription reaction (Enzo® BioArray™ HighYield™ RNA Transcript Labeling Kit, ENZO). After quantification of the purified cRNA (RNeasy Mini Kit, Qiagen), 15 ug are fragmented by alkaline treatment (200 mM Tris-acetate, pH 8.2, 500 mM potassium acetate, 150 mM magnesium acetate) and added to the hybridization cocktail sufficient for 5 hybridizations on standard GeneChip microarrays. Before expression profiling Test3 Probe Arrays (Affymetrix) are chosen for monitoring of

the integrity of the cRNA. Only labeled cRNA-cocktails which showed a ratio of the messured intensity of the 3' to the 5' end of the GAPDH gene less than 3.0 are selected for subsequent hybridization on HG-U133 probe arrays (Affymetrix). Washing and staining the Probe arrays is performed as described (siehe Affymetrix-Original-Literatur (LOCKHART und LIPSHUTZ). The Affymetrix software (Microarray Suite, Version 4.0.1) extracted fluorescence intensities from each element on the arrays as detected by confocal laser scanning according to the manufacturers recommendations.

Table 1

#	affy	HUGO name	zres	pres	qres	Map location
1	243479_at		3.643716	0.00026873	0.8912802	
2	240152_at		3.617023	0.00029801	0.8912802	
3	233098_s_at	DKFZp761N1814	3.609356	0.00030696	0.8912802	
4	208513_at	FOXB1	3.599972	0.00031825	0.8912802	15q21-q26
5	232659_at		3.545863	0.00039133	0.8912802	
6	206548_at	FLJ23556	3.526279	0.00042144	0.8912802	10q25.3
7	227384_s_at		3.455434	0.00054941	0.8912802	
8	231007_at		3.442766	0.0005758	0.8912802	
9	223543_at	KIAA1444	3.374117	0.00074053	0.8912802	Xq28
10	220553_s_at	FLJ20666	3.336705	0.00084778	0.8912802	14q21.1
11	238784_at	FLJ32949	3.300335	0.00096569	0.8912802	12q14.1
12	230030_at	HS6ST2	3.299677	0.00096796	0.8912802	Xq26.2
13	223567_at	SEMA6B	3.252241	0.00114499	0.8912802	19p13.3
14	235599_at		3.217945	0.00129112	0.8912802	
15	232528_at		3.215265	0.00130324	0.8912802	
16	243252_at		3.192369	0.00141111	0.8912802	
17	217163_at		3.191142	0.00141712	0.8912802	
18	218664_at	CGI-63	-3.18873	0.00142899	0.8912802	1pter-p22.3
19	243003_at		3.166166	0.00154463	0.8912802	
20	239811_at		3.140374	0.00168732	0.8912802	
21	207170_s_at	DKFZP586A011	3.138007	0.00170101	0.8912802	12q13.12
22	233120_at		3.131825	0.00173723	0.8912802	
23	230387_at		3.130418	0.00174558	0.8912802	
24	212939_at	COL6A1	3.125158	0.0017771	0.8912802	21q22.3
25	226408_at	TEAD2	3.122448	0.00179354	0.8912802	19q13.3
26	242457_at		3.102483	0.00191905	0.8912802	
27	239385_at	TFG	3.096979	0.00195504	0.8912802	3q11-q12
28	244414_at		3.096788	0.0019563	0.8912802	
29	244702_at		3.094677	0.00197028	0.8912802	
30	229879_at		3.09192	0.00198866	0.8912802	
31	232324_x_at		3.086977	0.00202203	0.8912802	

32	241309_at		3.08162	0.00205878	0.8912802	
33	237519_at		3.073104	0.00211844	0.8912802	
34	234279_at		3.059724	0.00221541	0.8912802	
35	235563_at		3.03083	0.00243882	0.8912802	
36	230432_at		3.030547	0.00244111	0.8912802	
37	237943_at		3.026426	0.00247464	0.8912802	
38	230401_at		3.017427	0.00254931	0.8912802	
39	228455_at		3.014548	0.00257362	0.8912802	
40	204086_at	PRAME	2.999359	0.00270549	0.8912802	22q11.22
41	209074_s_at	TU3A	-2.99885	0.00271001	0.8912802	3p21.1
42	243356_at		2.995792	0.00273733	0.8912802	
43	228392_at		2.992637	0.00276578	0.8912802	
44	214753_at		2.989486	0.00279447	0.8912802	
45	212210_at	DKFZP586J0619	-2.984664	0.0028389	0.8912802	7p22.3
46	222282_at		2.980831	0.00287468	0.8912802	
47	240733_at		2.977814	0.00290312	0.8912802	
48	244290_at		2.973458	0.00294465	0.8912802	
49	232615_at		2.971496	0.00296353	0.8912802	
50	227383_at		2.962167	0.00305482	0.8912802	

Table 2

#	affy	HUGO name	zres	pres	qres	Map location
1	206548_at	FLJ23556	3.653263	0.00025893	0.8975892	10q25.3
2	240152_at		3.637236	0.00027558	0.8975892	
3	243479_at		3.563259	0.00036628	0.8975892	
4	227384_s_at		3.528296	0.00041824	0.8975892	
5	231007_at		3.491441	0.00048042	0.8975892	
6	232659_at		3.459668	0.00054084	0.8975892	
7	233098_s_at	DKFZp761N1814	3.435508	0.00059144	0.8975892	
8	238784_at	FLJ32949	3.434273	0.00059415	0.8975892	12q14.1
9	228392_at		3.392899	0.00069157	0.8975892	

10	220553_s_at	FLJ20666	3.378526	0.00072876	0.8975892	14q21.1
11	223543_at	KIAA1444	3.373988	0.00074088	0.8975892	Xq28
12	244414_at		3.359689	0.0007803	0.8975892	
13	237519_at		3.343881	0.00082615	0.8975892	
14	233120_at		3.315484	0.00091485	0.8975892	
15	212939_at	COL6A1	3.314391	0.00091843	0.8975892	21q22.3
16	242457_at		3.287439	0.00101103	0.8975892	
17	237943_at		3.264744	0.00109563	0.8975892	
18	230387_at		3.224425	0.00126226	0.8975892	
19	241309_at		3.192535	0.0014103	0.8975892	
20	232851_at		3.186605	0.00143953	0.8975892	
21	232528_at		3.185757	0.00144376	0.8975892	
22	243003_at		3.185516	0.00144496	0.8975892	
23	244290_at		3.18345	0.00145531	0.8975892	
24	205470_s_at	KLK11	-3.168772	0.00153084	0.8975892	19q13.3-q13.4
25	200686_s_at	SFRS11	3.164092	0.00155567	0.8975892	1p31
26	217163_at		3.15324	0.00161469	0.8975892	
27	233352_at		3.147222	0.0016483	0.8975892	
28	232324_x_at		3.141528	0.00168069	0.8975892	
29	204751_x_at	DSC2	3.13979	0.00169069	0.8975892	18q12.1
30	231348_s_at	DAT1	3.110996	0.00186458	0.8975892	12p12.3
31	239725_at		3.086411	0.00202589	0.8975892	
32	234119_at		3.081957	0.00205645	0.8975892	
33	239811_at		3.075844	0.00209908	0.8975892	
34	222736_s_at	FLJ10493	3.068783	0.00214932	0.8975892	9q31.2
35	228455_at		3.063715	0.00218607	0.8975892	
36	237180_at		3.062217	0.00219704	0.8975892	
37	240146_at		3.05469	0.00225293	0.8975892	
38	230432_at		3.032986	0.00242147	0.8975892	
39	243252_at		3.028124	0.00246077	0.8975892	
40	244702_at		3.024874	0.00248737	0.8975892	
41	211185_s_at	FLJ14753	3.022503	0.00250694	0.8975892	9q22.31
42	221899_at	CG005	3.021914	0.00251182	0.8975892	13q12-q13
43	234279_at		3.019776	0.00252961	0.8975892	

44	205894_at	ARSE	-3.015471	0.0025658	0.8975892	Xp22.3
45	235521_at	HOXA3	3.012118	0.00259432	0.8975892	7p15-p14
46	221907_at	FLJ40452	-3.010951	0.00260431	0.8975892	14q32.33
47	215183_at		-3.009188	0.00261947	0.8975892	
48	218709_s_at	C20orf9	-3.008784	0.00262296	0.8975892	
49	210537_s_at	TADA2L	3.003556	0.00266844	0.8975892	17q12-q21
50	241858_at		3.000611	0.00269439	0.8975892	

Explanatory Table 1A

#	affy							
1	243479_at	Hs.35758.0	H69055	Hs.35758.0_RC	GenBank	Hs.35758		
2	240152_at	Hs.127922.0	BF792954	Hs.127922.0_RC	GenBank	Hs.127922		
3	233098_s_at	Hs.283780.0	AL353947.1	Hs.283780.0	GenBank	Hs.283780	55360	
4	208513_at	Hs.247756.0	NM_012182.1	g11386194	RefSeq	Hs.247756	27023	NM_012182; forkhead box B1
5	232659_at	Hs.287480.0	AU146864	Hs.287480.0	GenBank	Hs.287480		
6	206548_at	Hs.214039.0	NM_024880.1	g13376321	RefSeq	Hs.214039	79938	NM_024880; hypothetical protein FLJ23556
7	227384_s_at	Hs.36475.0	AW340595	Hs.36475.0.S1	GenBank	Hs.36475		
8	231007_at	Hs.164129.0	AI565054	Hs.164129.0.A1	GenBank	Hs.444693		
9	223543_at	Hs.92732.0	BC002606.1	g12803550	GenBank	Hs.92732	57595	NM_032512; LU1 protein
10	220553_s_at	Hs.250477.0	NM_018333.1	g8922887	RefSeq	Hs.274337	55015	NM_017922; hypothetical protein FLJ20666 NM_018333; hypothetical protein FLJ20666
11	238784_at	Hs.125472.0	AI039361	Hs.125472.0_RC	GenBank	Hs.125472	283417	NM_173812; hypothetical protein FLJ32949
12	230030_at	Hs.82302.0	AI767756	Hs.82302.0.A1	GenBank	Hs.82302	90161	NM_147174; heparan sulfate 6-O-sulfotransferase 2 NM_147175; heparan sulfate 6-O-sulfotransferase 2 isoform S
13	223567_at	Hs.148932.1	AB022433.1	g12081906	GenBank	Hs.148932	10501	NM_020241; semaphorin 6B isoform 1 precursor NM_032108; semaphorin 6B isoform 3 precursor NM_133327; semaphorin 6B isoform 2 precursor
14	235599_at	Hs.125346.0	AW105723	Hs.125346.0_RC	GenBank	Hs.125346		

15	232528_at	Hs.270124.0	AI338705	Hs.270124.0	GenBank	Hs.270124		
16	243252_at	Hs.177588.0	AA173465	Hs.177588.0	GenBank	Hs.439082		
17	217163_at	Hs.247938.0	X63118	Hs.247938.0.S1	GenBank			
18	218664_at	Hs.19513.0	NM_016011.1	g7705776	RefSeq	Hs.19513	51102	NM_016011; nuclear receptor-binding factor 1
19	243003_at	Hs.69606.0	AV702197	Hs.69606.0_RC	GenBank	Hs.69606		
20	239811_at	Hs.129037.0	BF954306	Hs.129037.0_RC	GenBank	Hs.129037		
21	207170_s_at	Hs.75884.0	NM_015416.1	g7661659	RefSeq	Hs.75884	25875	NM_015416; cervical cancer 1 protooncogene protein p40
22	233120_at	Hs.287602.0	AK023907.1	Hs.287602.0_RC	GenBank			
23	230387_at	Hs.48948.0	AL038450	Hs.48948.0.A1	GenBank	Hs.48948		
24	212939_at	Hs.25459.0	M20776.1	Hs.25459.0.A1	GenBank	Hs.108885	1291	NM_001848; collagen, type VI, alpha 1 precursor
25	226408_at	Hs.153053.1	AA905942	Hs.153053.1.A1	GenBank	Hs.166556	8463	NM_003598; TEA domain family member 2
26	242457_at	Hs.257396.0	AW451107	Hs.257396.0.A1	GenBank	Hs.69504		
27	239385_at	Hs.142230.0	AI150613	Hs.142230.0_RC	GenBank	Hs.250897	10342	NM_006070; TRK-fused gene
28	244414_at	Hs.222120.0	AI148006	Hs.222120.0_RC	GenBank	Hs.222120		
29	244702_at	Hs.195381.0	AI654208	Hs.195381.0_RC	GenBank	Hs.195381		
30	229879_at	Hs.29419.0	BF059124	Hs.29419.0_RC	GenBank	Hs.396842		
31	232324_x_at	Hs.302480.0	AK001092.1	Hs.302480.0.S1	GenBank			
32	241309_at	Hs.147254.0	BE466813	Hs.147254.0_RC	GenBank	Hs.147254		
33	237519_at	Hs.148609.0	BE[illegible]783	Hs.148609.0_RC	GenBank	Hs.148609		

34	234279_at	Hs.306343.0	AL117453.1	Hs.306343.0.S1	GenBank	Hs.306343		
35	235563_at	Hs.256862.0	BG2[illegible]68	Hs.256862.0.A1	GenBank	Hs.288660		
36	230432_at	Hs.14535.0	AI733124	Hs.14535.0.S1	GenBank	Hs.14535		
37	237943_at	Hs.246358.0	AI820802	Hs.246358.0.A1	GenBank	Hs.246358		
38	230401_at	Hs.125109.0	BF197705	Hs.125109.0.A1	GenBank	Hs.125109		
39	228455_at	Hs.126465.0	AI092824	Hs.126465.0_RC	GenBank	Hs.126465		
40	204086_at	Hs.30743.0	NM_006115.1	g5174640	RefSeq	Hs.30743	23532	NM_006115; preferentially expressed antigen in melanoma
41	209074_s_at	Hs.8022.1	AL050264.1	g4886486	GenBank	Hs.8022	11170	NM_007177; downregulated in renal cell carcinoma
42	243356_at	Hs.233461.0	N34972	Hs.233461.0.A1	GenBank	Hs.233461		
43	228392_at	Hs.282588.0	BF508739	Hs.282588.0	GenBank	Hs.282588		
44	214753_at	Hs.110630.0	AW084068	Hs.110630.0	GenBank	Hs.110630		
45	212210_at	Hs.112184.0	AB037861.1	Hs.112184.0	GenBank	Hs.112184	26173	
46	222282_at	Hs.294014.0	AV761453	Hs.294014.0.A1	GenBank	Hs.294014		
47	240733_at	Hs.118394.0	W92005	Hs.118394.0.A1	GenBank	Hs.118394		
48	244290_at	Hs.252627.0	AW293174	Hs.252627.0_RC	GenBank	Hs.444018		
49	232615_at	Hs.163986.0	AA632758	Hs.163986.0.A1	GenBank	Hs.163986		
50	227383_at	Hs.36475.0	AW340595	Hs.36475.0.S1	GenBank	Hs.36475		

Explanatory Table 2A

#	affy							
1	206548_at	Hs.214039.0	NM_024880.1	g13376321	RefSeq	Hs.214039	79938	NM_024880; hypothetical protein FLJ23556
2	240152_at	Hs.127922.0	BF792954	Hs.127922.0_RC	GenBank	Hs.127922		
3	243479_at	Hs.35758.0	H69055	Hs.35758.0_RC	GenBank	Hs.35758		
4	227384_s_at	Hs.36475.0	AW340595	Hs.36475.0.S1	GenBank	Hs.36475		
5	231007_at	Hs.164129.0	AI565054	Hs.164129.0.A1	GenBank	Hs.444693		
6	232659_at	Hs.287480.0	AU146864	Hs.287480.0	GenBank	Hs.287480		
7	233098_s_at	Hs.283780.0	AL353947.1	Hs.283780.0	GenBank	Hs.283780	55360	
8	238784_at	Hs.125472.0	AI039361	Hs.125472.0_RC	GenBank	Hs.125472	283417	NM_173812; hypothetical protein FLJ32949
9	228392_at	Hs.282588.0	BF508739	Hs.282588.0	GenBank	Hs.282588		
10	220553_s_at	Hs.250477.0	NM_018333.1	g8922887	RefSeq	Hs.274337	55015	NM_017922; hypothetical protein FLJ20666 NM_018333; hypothetical protein FLJ20666
11	223543_at	Hs.92732.0	BC002606.1	g12803550	GenBank	Hs.92732	57595	NM_032512; LU1 protein
12	244414_at	Hs.222120.0	AI148006	Hs.222120.0_RC	GenBank	Hs.222120		
13	237519_at	Hs.148609.0	BE463783	Hs.148609.0_RC	GenBank	Hs.148609		
14	233120_at	Hs.287602.0	AK023907.1	Hs.287602.0_RC	GenBank			
15	212939_at	Hs.25459.0	M20776.1	Hs.25459.0.A1	GenBank	Hs.108885	1291	NM_001848; collagen, type VI, alpha 1 precursor
16	242457_at	Hs.257396.0	AW451107	Hs.257396.0.A1	GenBank	Hs.69504		

17	237943_at	Hs.246358.0	AI820802	Hs.246358.0.A1	GenBank	Hs.246358		
18	230387_at	Hs.48948.0	AL038450	Hs.48948.0.A1	GenBank	Hs.48948		
19	241309_at	Hs.147254.0	BE466813	Hs.147254.0_RC	GenBank	Hs.147254		
20	232851_at	Hs.17992.0	AL162053.1	Hs.17992.0	GenBank	Hs.406787		
21	232528_at	Hs.270124.0	AI338705	Hs.270124.0	GenBank	Hs.270124		
22	243003_at	Hs.69606.0	AV702197	Hs.69606.0_RC	GenBank	Hs.69606		
23	244290_at	Hs.252627.0	AW293174	Hs.252627.0_RC	GenBank	Hs.444018		
24	205470_s_at	Hs.57771.0	NM_006853.1	g5803198	RefSeq	Hs.57771	11012	NM_006853; kallikrein 11 isoform 1 preproprotein NM_144947; kallikrein 11 isoform 2 precursor
25	200686_s_at	Hs.11482.0	NM_004768.1	g4759099	RefSeq	Hs.433581	9295	NM_004768; splicing factor p54
26	217163_at	Hs.247938.0	X63118	Hs.247938.0.S1	GenBank			
27	233352_at	Hs.287590.0	AK023753.1	Hs.287590.0	GenBank	Hs.287590		
28	232324_x_at	Hs.302480.0	AK001092.1	Hs.302480.0.S1	GenBank			
29	204751_x_at	Hs.239727.0	NM_004949.1	g13435365	RefSeq	Hs.239727	1824	NM_004949; desmocollin 2 isoform Dsc2b preproprotein NM_024422; desmocollin 2 isoform Dsc2a preproprotein
30	231348_s_at	Hs.301914.1	BF508869	Hs.301914.1.A1	GenBank	Hs.301914	55885	NM_018640; neuronal specific transcription factor DAT1
31	239725_at	Hs.16727.0	T90703	Hs.16727.0.A1	GenBank	Hs.16727		
32	234119_at	Hs.306484.0	AL157462.1	Hs.306484.0.A1	GenBank	Hs.306484		
33	239811_at	Hs.129037.0	BF954306	Hs.129037.0_RC	GenBank	Hs.129037		
34	222736_s_at	Hs.279610.0	BC000049.1	g12652608	GenBank	Hs.279610	55151	NM_018112; hypothetical protein FLJ10493

35	228455_at	Hs.126465.0	AI092824	Hs.126465.0_RC	GenBank	Hs.126465		
36	237180_at	Hs.119563.0	T97717	Hs.119563.0.A1	GenBank	Hs.119563		
37	240146_at	Hs.281196.0	AW418562	Hs.281196.0.A1	GenBank	Hs.281196		
38	230432_at	Hs.14535.0	AI733124	Hs.14535.0.S1	GenBank	Hs.14535		
39	243252_at	Hs.177588.0	AA173465	Hs.177588.0	GenBank	Hs.439082		
40	244702_at	Hs.195381.0	AI654208	Hs.195381.0_RC	GenBank	Hs.195381		
41	211185_s_at	Hs.13453.2	AF130099.1	g11493501	GenBank	Hs.13453	84641	NM_032558; hypothetical protein FLJ14753
42	221899_at	Hs.23518.1	AI809961	Hs.23518.1_RC	GenBank	Hs.23518	10443	NM_014887; hypothetical protein from BCRA2 region
43	234279_at	Hs.306343.0	AL117453.1	Hs.306343.0.S1	GenBank	Hs.306343		
44	205894_at	Hs.74131.0	NM_000047.1	g4502240	RefSeq	Hs.74131	415	NM_000047; arylsulfatase E precursor
45	235521_at	Hs.222446.0	AW137982	Hs.222446.0.A1	GenBank	Hs.248074	3200	NM_030661; homeobox A3 protein isoform a NM_153631; homeobox A3 protein isoform a NM_153632; homeobox A3 protein isoform b
46	221907_at	Hs.81920.0	AI679213	Hs.81920.0_RC	GenBank	Hs.339834	115708	NM_152307; hypothetical protein FLJ40452
47	215183_at	Hs.284192.0	AF090886.1	Hs.284192.0.A1	GenBank	Hs.406792		
48	218709_s_at	Hs.24994.0	NM_016004.1	g7705768	RefSeq	Hs.24994	51098	NM_016004; chromosome 20 open reading frame 9
49	210537_s_at	Hs.125156.1	BC001172.1	g12654666	GenBank	Hs.125156	6871	NM_001488; transcriptional adaptor 2-like isoform a NM_133439; transcriptional adaptor 2-like isoform b
50	241858_at	Hs.47122.0	AA707390	Hs.47122.0.A1	GenBank	Hs.47122		